Chinese L2 Learners' Depth of Vocabulary Knowledge and Its Role in Reading Comprehension

Introduction

Vocabulary was once regarded as a neglected aspect of language teaching and learning (Meara, 1980). Three decades later, it has arguably become one of the most extensively researched areas in second language (L2) acquisition, instruction, and assessment (Schmitt, 2010). Yet, studies on L2 vocabulary knowledge showed a clear imbalance with respect to its multidimensionality. For example, a strong knowledge base has been established about the dimension of vocabulary size or breadth (i.e., how many words one knows), including its assessment (Nation, 1990; Read, 2002) and its critical import for the development of L2 abilities, notably reading comprehension (Grabe, 2009; Jeon & Yamashita, 2014; Koda, 2005; Milton, 2013; Nation, 2001). However, far less is known about vocabulary depth or the dimension of vocabulary knowledge that pertains to the quality of the knowledge that one has about words (Schmitt, 2014). Existing studies that involved vocabulary depth focused primarily on English as a Second/Foreign language (e.g., Li & Kirby, 2015; Qian, 1999, 2002; Read, 1993, 1998; Zhang, 2012). Little is known about the vocabulary depth of L2 learners of Chinese, the focal language of the present study; and the impact it has on their Chinese reading development. To this end, using a Chinese Word Associates Test that the authors developed, this study examined Chinese L2 learners' depth of vocabulary knowledge and its role in their reading comprehension.

Review of Literature

Vocabulary Depth and Word Association

Among the various conceptualizations of what it means to know a word, the best-known one is perhaps the differentiation between size/breadth and depth dimensions of vocabulary knowledge

(Anderson & Freebody, 1981; Schmitt, 2014; Wesche & Paribakht, 1996). While there is a consensus that vocabulary depth concerns the quality of the knowledge that a learner has about words, what exactly constitutes quality or what specific aspects of knowledge depth entails has been conceptualized and discussed in diverse ways (e.g., Henriksen, 1999; Milton, 2009; Read, 2004; Richards, 1976; Schmitt, 2014; Wesche & Paribakht, 1996). In an oft-cited paper, Read (2004, pp. 211-212), for example, distinguished between three meanings of depth, including precision of meaning ("the difference between having a limited unclear idea of what a word means and having much more specific knowledge of its meaning"), comprehensive word knowledge ("knowing the semantic feature of a word and its orthographic, phonological, morphological, syntactic, collocational and pragmatic characteristics"), and network knowledge ("the incorporation of the word into its related words in the schemata, and the ability to distinguish its meaning and use from related words").

With a focus on the network knowledge aspect, Read (1993, 1998) developed a Word Associates Test (WAT) for English as a Second Language (ESL) learners based on the concept of word association. As Read (2004) argued, "as a learner's vocabulary size increases, newly acquired words need to be accommodated within a network of already known words, and some restructuring of the network may be needed as a result." "This means that depth can be understood in terms of learners' developing ability to distinguish semantically related words and, more generally, their knowledge of the various ways in which individual words are linked to each other." (p. 219) Ever since Read (1993, 1998) developed the prototypes of WAT, various other forms of the test have been developed, depending on specific design features, such as number of choices and presentation format (e.g., Greidanus & Nienhuis, 2001; Qian & Schedl, 2004; Schmitt, Ng, & Garras, 2011; Schoonen & Verhallen, 2008). Typically, in a WAT item, a

target word is followed by six or eight other words, half of which are associated with the target word (i.e., associates) and the other half are not (i.e., distractors). The associates have two primary types of relationship with a target word: paradigmatic and syntagmatic. The former refers to an associate from the same word class and performs the same grammatical function as the target word in a sentence (e.g., a synonym), whereas the latter pertains to an associate with a sequential relationship to the target word in a sentence and is usually a word from a different word class (e.g., a collocate).

Validation studies have shown that tests following the word associates format are reliable and valid in assessing learners' vocabulary depth (Greidanus, Boggards, van der Linden, Nienhuis, & de Wolf, 2004; Read, 1993, 1998; Schmitt et al., 2011; Schoonen & Verhallen, 2008). Various forms of the WAT have also been used to investigate the organization of words in L2 learners' mental lexicon and the development of different types of association knowledge (and the role of vocabulary depth in language proficiency development, including reading comprehension, which is reviewed in the next section). A common way of exploring L2 learners' semantic organization of words, which follows the L1 literature (Nelson, 1977), is to ask students to give free associations for a selected set of stimulus words (Fitzpatrick, 2013). Learners' responses are then categorized in accordance to different types of association relationships, sometimes in comparison to the responses in their native language (L1) (e.g., Nissen & Henriksen, 2006; Wolter, 2001). However, vocabulary scholars have expressed concerns about free association tasks, despite the fact that such tasks have been shown to provide insightful information about learners' lexical organization (e.g., paradigmatic-syntagmatic shift; Wolter, 2001). For example, learners' responses could be "diverse and unstable" (Read, 1993, p. 358); thus, compared to the controlled form of word association such as in the case of Read's

(1993, 1998) WAT, the free association format may not be able to provide a reliable estimate of the network knowledge that learners have. In addition, scoring of learners' vocabulary production can be a challenging, if not impossible, approach when documenting individual differences among learners (Henriksen, 2008; Fitzpatrick, Playfoot, Wray, & Wright, 2013)

The aforementioned limitations of free association were perhaps a reason that in studies where learners' vocabulary depth needed to be measured, the controlled format of WAT was often used (e.g., Greidahus & Nienhuis, 2001; Horiba, 2012; Zhang, 2012). In Greidahus and Nienhuis's (2001) study on Dutch-speaking university learners of French, for example, three types of association relationships, including paradigmatic, syntagmatic, and analytic (defining characteristics of a word), were distinguished. The target words included five groups of nouns, adjectives, and verbs, with each group sampled from a distinct frequency band. Among other findings, higher proficiency learners performed significantly better than lower proficiency learners. Learners' performance on items with higher-frequency target words was also significantly better than that on those with target words having a lower frequency. In addition, for both groups of learners, the scores for both paradigmatic association and analytic association were significantly higher than those for syntagmatic association. Horiba (2012) found Korean-speaking learners of Japanese performed significantly better on more frequent target words than on less frequent words; yet, such a word frequency effect was not found significant for Chinese-speaking learners of Japanese. Significant score differences among the three types of association relationships examined in Greidanus and Nienhus (2001) were only observed for Chinese-speaking learners, with paradigmatic association showing the best performance.

Overall, the WAT literature did not seem to have produced consistent findings on the effect of word frequency on test scores. While learners tended to show significantly better

performance for paradigmatic association than syntagmatic association, which appears to be aligned with the predominance of paradigmatic associates revealed in previous studies on L2 learners' free association, it remains unknown how WAT performance may be a function of the joint influence of type of association relationships and word class of target words. Nissen and Henriksen (2006) found from a free association task that the proportional representation of different types of association relationships in Danish-speaking English L2 learners' responses was actually moderated by the form classes of the stimulus words. However, no previous studies seemed to have addressed the effect of word class in the WAT literature, even though some of them did include words of different form classes in their tests (e.g., Greidanus & Nienhus, 2001). It was thus the interest of the present study to further address the impact of word frequency, type of association relationships, and word class of target words on WAT performance with a focus on L2 learners of Chinese.

Vocabulary Depth in Reading Comprehension

Logically, as words are the building block of language, one needs to know the meanings of the words in a text in order to comprehend it (i.e., the instrumentalist hypothesis; Anderson & Freebody, 1981). Thus, the more words one knows, the better one would be at reading comprehension. There is thus no wonder that much research attention in the L2 vocabulary and reading literature has been paid to the critical import of vocabulary size (Grabe, 2009; Jeon & Yamashita, 2014; Koda, 2005; Milton, 2013). It was found that a lexical coverage of at least 98% is needed for one to have adequate and unassisted comprehension of a text (Hu & Nation, 2000; Nation, 2006). Vocabulary size has also been found to have a strong correlational relationship with reading comprehension, with correlations as high as .80 (Jeon & Yamashita, 2014; Milton, 2013; Qian, 1999, 2002; Zhang, 2012).

Compared to vocabulary size, how vocabulary depth contributes to reading comprehension has been much less well examined (Grabe, 2009; Jeon & Yamashita, 2014). As learners' proficiency increases, those who know more words often tend to know those words better, which often leads to concerns about whether vocabulary size and depth are indeed distinct aspects of vocabulary knowledge on the one hand (Vermeer, 2001) and how they are relatively important to L2 reading comprehension on the other (Horiba, 2012; Li & Kirby, 2015; Qian, 1999, 2002).

In a study of ESL learners in Canadian universities, Qian (1998) found that learners' vocabulary depth, which was measured with Read's (1993) WAT, explained about 11% of the additional variance in their reading comprehension, after controlling for vocabulary size, which was measured with the Vocabulary Levels Test (VLT) (Nation, 1990). In addition, over and above the VLT, the WAT was also a significant, unique predictor of reading comprehension. However, the unique proportion of the variance in reading comprehension that was explained by the WAT (about 3%) was far less than that by the VLT. These findings suggested that vocabulary size and depth are distinct aspects of vocabulary knowledge based on their unique predictive effects, and that vocabulary depth tended to be a more important contributor to reading comprehension. Yet, in Horiba's (2012) study on Chinese- and Korean-speaking learners of L2 Japanese, no unique and significant effect of vocabulary depth on reading comprehension surfaced when the effect of vocabulary size was considered.

Li and Kirby (2015) further suggested that the extent to which vocabulary size and depth are relatively contributive to reading comprehension may actually depend on the comprehension tasks themselves, which may place different levels of processing demands on readers. In their study on Chinese-speaking adolescent readers of English as a Foreign Language in China, Li and

Kirby (2015) measured reading comprehension with both a standardized multiple-choice reading comprehension test and a summary writing task. When the former was the criterion variable of regression analysis, vocabulary depth significantly explained about 5% of the additional variance in reading comprehension when it was entered before vocabulary breadth but after other non-vocabulary covariates. Over and above depth and other related variables, size also had a significant unique effect on reading comprehension and explained about 9% of its variance. When the order of entry for breadth and depth was switched, the result was different. Specifically, vocabulary breath explained a significant and unique amount of variance (about 13%) in reading comprehension. Over and above breadth, however, depth no longer significantly predicted reading comprehension and the variance explained was negligible. When the summary writing task was the criterion variable, vocabulary depth explained about 5% of the unique variance in comprehension; over and above depth, size no longer significantly predicted summary writing. Controlling for size, which had a significant effect on reading comprehension (about 2% unique variance explained), however, depth still significantly predicted summary writing (about 4% of variance explained). Taken together, these findings lent clear support for task effects on the relative contributions of vocabulary size and depth to reading comprehension.

Li and Kirby's (2015) findings seem to be in line with the heuristic proposed by the RAND Reading Study Group (2002) on the interaction between reader and non-reader variables (e.g., text and activity/task) in textual comprehension as well as the findings of many L1 studies (e.g., Eason, Goldberg, Young, Geist, & Cutting, 2012; Hamilton, Freed, & Long, 2013). Yet, how the relative contributions of sub-skills, including different types of vocabulary knowledge, to reading comprehension may vary as a function of the texts and/or tasks has received little attention in the L2 reading literature. Jeon and Yamashita's (2014) meta-analysis identified a

number of factors that tended to affect how different types of linguistic knowledge, skills, and strategies contributed to L2 passage comprehension, including differential definitions, operationalizations, and measurements of some independent variables. Although it makes an important contribution, their meta-analysis did not consider potential moderations that were induced by variables at the level of reading tasks and/or textual properties. Thus, the study reported here investigated the relative contributions of vocabulary size and depth to L2 Chinese reading comprehension on the one hand and how these contributions to L2 reading comprehension may vary as a function of texts and reading tasks on the other. Specifically, it addressed the following three questions:

1. Do word frequency, word class, and type of association relationship have an impact on Chinese L2 learners' word association performance?

2. Is vocabulary depth a unique predictor of Chinese L2 learners' reading comprehension over and above vocabulary size?

3. Do the relative contributions of vocabulary size and depth vary as a function of texts and reading tasks?

Methods

Participants

The participants were 21 students studying abroad at a university in China. None of them were heritage learners of Chinese. Their native languages varied, including, for example, English, Russian, and Korean. They included two males and 19 females with an average age of about 22.5 years. A background survey revealed that 16 of them (about 76%) had studied Chinese in their home country for about two years or more, and 18 of them (about 86%) had studied in China for about a year. The only student who had not learned any Chinese in her own country had studied

in China for more than three years. Most of the participants chose to study in China in order to enhance their Chinese proficiency, and did not have a specific major at the host university.

Measures

In addition to a survey that revealed the aforementioned background about the learners, the participants also completed a Chinese WAT test (hereafter, WAT-C), an orthographic processing task, a picture selection task, a grammaticality judgement task, and two multiple-choice passage comprehension tasks.

WAT-C. The WAT-C is a 20-item task which measured learners' depth of vocabulary knowledge. Target words included 10 adjectives and 10 verbs, each of which was followed by two boxes of three words with three of the six words being associates and the other three being distractors. As in Read's (1998) English WAT, words in the left box were intended to tap paradigmatic relationships and words in the right box, syntagmatic relationships. Below is an example of an adjective item and a verb item. To illustrate, 好看 is an adjective that means *good-looking* or *pretty*, and has 漂亮 (*pretty*) and 美丽 (*pretty*/*beautiful*) as their synonymic associates; it also has 图片 (*picture*) as its collocate, but not 主意(*idea*/*thought*) and 工作 (*job*). 忘记 is a verb that means *forget*, and should have 想起 (*to recall* or *to come back to memory*), but not 觉得 (*to think/believe/gather*) and 发出 (*to send out* or *to issue*), as its paradigmatic associate; and 事情 (*things*) and 历史 (history), but not 星期 (*week*), as its syntagmatic associates.

好看

(1) 漂亮 (2) 幸运 (3) 美丽	(4) 主意 (5) 图片 (6) 工作

忘记

(1) 觉得 (2) 想起 (3) 发出	(4) 事情 (5) 星期 (6) 历史

All words, including the 20 target words and the 120 choice words, were selected from the *Graded Chinese Syllables, Characters and Words for the Application of Teaching Chinese to Speakers of Other Languages* (hereafter, GCSCW-TCSOL) (State Language Commission, 2011). The graded words consist of three levels that cover 11,092 words, including 2,245 for Level 1, 3,211 for Level 2, and 5,636 for Level 3. Level 1 is the lowest level and includes highly frequent words for entry level learners of Chinese. It was also the level from which the 140 WAT-C words were selected, which were all two-character compound words. Table 1 shows the average raw and log10 frequencies of the target words and their choice words in accordance to the two word classes (i.e., adjectives and verbs). The raw frequency came from a wordlist generated by the State Language Commission (available on http://www.cncorpus.org/resources.aspx) based on a corpus of 200 million Chinese characters. The log10 frequency of the 10 target adjectives was significantly lower than that of the 10 target verbs, $t(18) = -2.625$, $p = .017$. ANOVA with the log10 frequency of the choice words as the dependent variable revealed a significant main effect of word class, $F(1, 18) = 7.758$, $p = .012$, which indicated that, disregarding types of association, the frequency of choice words for target verbs was significantly higher than that of those choice words for target adjectives. No main effect of association relationships, however, was found, $F(1, 18) = 1.255$, $p = .277$, which indicated that, disregarding word class, the frequency was similar between the choice words for paradigmatic and syntagmatic relationships.

Insert Table 1 about here

When the WAT-C was administered, learners were informed that each box had one or two words that had a close relationship with the target word, and the total number of words selected should be three. To make it possible to compare learners' scores for syntagmatic and paradigmatic associates, the scoring method used by Greidahus and colleagues (Greidahus & Nienhuis, 2001; Greidahus et al., 2004) was adopted; this system gives credit for both the selection of associates as well as the non-selection of distractors. In other words, learners were awarded one point if they chose an associate (either syntagmatic or paradigmatic) and if they did not choose a distractor. A score of zero would be awarded if they did not provide a response. With this scoring method, the range of score for syntagmatic as well as paradigmatic relationships for a WAT-C item would be 0-3 with the range for the whole item being 0-6.

Orthographic Processing. Given that the WAT-C was administered in print, to control for any influence of the learners' orthographic processing skills on their WAT-C performance, an Orthographic Choice task was administered. The task had 20 pairs of two-character words. In each pair, one word had the correct orthographic form (e.g., 错误[*error*] and 电脑[*computer*]), whereas the other word had a character that orthographically resembled the correct character but in which one component of the character was wrong (e.g., 借误 and 电恼). The learners were asked to circle the correct word for each pair. The maximum score was 20.

Vocabulary Size. A picture selection task was administered as a measure of the learners' vocabulary size. It included five single-character and 25 two-character words of various frequency levels based on the Modern Chinese Frequency Dictionary (Beijing Language Institute, 1986). Each word was followed by four pictures. The learners were asked to circle the picture that best represented the meaning of a target word. One point was awarded for each correct picture selection, with the total score being 30. Note that this test was not intended to provide an

estimate for the actual size of learners' vocabulary knowledge; rather, it was to designed to help identify any individual difference in the number of words that the learners could possibly know so as to predict their reading comprehension in conjunction with vocabulary depth.

Grammatical Knowledge. Grammatical knowledge is as a critical factor in L2 reading comprehension (Grabe, 2009; Koda, 2005; Urquhart & Weir, 1998; Zhang, 2012). To obtain a more accurate understanding of the unique effect of vocabulary depth on reading comprehension on the one hand and the relative contributions of vocabulary depth and size on the other, learners' grammatical knowledge was measured with a grammaticality judgment task and included as a covariate when reading comprehension was predicted by vocabulary knowledge in the regression analysis. The task included 10 pairs of lexically simple sentences. The two sentences in each pair were lexically different but tapped the same grammatical structure, with one sentence being grammatically appropriate and the other not. Learners were asked to circle Yes or No to indicate whether a sentence was grammatically appropriate. The aspects of grammar covered included, for example, aspectual markers, comparative structures, 把(*ba*) / 被(*bei*) structures, discourse markers like 吧, place and time adverbials, measure words, etc. To give an example, *可乐喝了被妹妹 (literally, **Coke drunk* bei *younger sister*) and 水果被妈妈吃了(literally, *Fruits* bei *mom eaten*) are a pair of sentences focusing on the *bei* structure. The first is ill-formed and the second one has the structure appropriately used, because *bei* + somebody (*[done] by someone*) should be placed before a verb to indicate that an action has been performed by the person. The maximum score was 20.

Reading Comprehension. Two passage comprehension tasks were administered, including a long passage comprehension task and a short passage comprehension task. The short passage task consisted of 15 narrative passages; each included only a few sentences (21-91

characters, with an average of 49 characters) and was followed by an inferential question, which asked the learners to integrate multiple pieces of information to infer about a message that was not explicitly stated in a passage. For example, 我去年春节去过那个小镇，今年再去的时候，经过那条街道，我几乎不认识了。(*I was in that small town during the Spring Festival last year. This year when I was there and passing that street, I could barely recognize it*.). Based on this passage, learners were asked to select an appropriate inference about the town: (1) 春节很热闹 (*Spring Festival was exciting*), (2) 变化很大 (*There were big changes*), and (3) 人很热情 (*People were very nice*). The answer should be (2). The maximum score was 15.

The long passage task included four long stories. The length of the four long passages ranged from 264 to 661 characters, including punctuation, with an average of 432 characters. Each long passage was followed by five multiple-choice questions. Altogether there were 20 questions, most of which were literal questions that measured the learners' identification of explicit information that was presented in the stories; a small number of questions also tested learners' ability to resolve co-referential relationships and get the main idea of a story. The maximum score was 20.

Five native speakers who all had experience of teaching Chinese as a Second/Foreign Language rated the lexical and grammatical complexity of all the short and long passages for intermediate/advanced learners of Chinese on a 7-point Likert scale with 1 indicating very simple and 7 very complex. The 15 short passages received an average rating of 1.65 (range 1-3) for lexical complexity and 1.84 (range 1-4) for grammatical complexity; the four long passages received an average rating of 4.8 (range 4-6) for lexical complexity and 4.35 (range 3-5) for grammatical complexity. Thus, the long passages were not only longer but also lexically much richer and grammatically more complex than the short passages.

Results

Descriptive Statistics and Reliabilities

Table 2 shows the learners' performance on all the tasks and task reliability (*n* = number of items). Overall, the tasks all had very good internal consistency reliability, except the Orthographic Processing and Grammaticality Judgment tasks, whose reliability appeared to be low but acceptable.

Insert Table 2 about here

Word Frequency Effect on WAT-C Performance

To examine the effect of word frequency on learners' performance on the WAT-C, correlations were calculated between different (log10) frequency indexes for the WAT-C items and the average performance of the 21 learners for each item. The frequency indexes included the frequency of target words only; the average frequency of six choice words for each target word, and the average frequency of all seven words in a WAT-C item. The correlations were all positive, but none was significant: $r = .105$, $p = .659$; $r = .287$, $p = .219$; and $r = .264$, $p = .261$, respectively, for the three frequency indexes.

Word Class, Association Type, and WAT-C Performance

To further address the relationship between word frequency and learners' performance on the WAT-C, a two-way ANCOVA was conducted with word class (i.e., adjectives vs. verbs) and type of association relationships (i.e., paradigmatic vs. syntagmatic) as the independent variables, WAT-C performance as the dependent variable, and orthographic processing as the covariate. Orthographic processing explained about 19.5% of the variance in the dependent variable, $F(1,$

19) = 4.589, p = .045. After adjusting for the influence of the covariate, there was no significant main effect of word type, $F(1, 19) = .708$, $p = .411$, which indicated that, disregarding association relationships, learners' performance on adjectives (M = 25.643, SE = .603) and verbs (M = 25.881, SE = .788) was not significantly different. No significant main effect was found for association relationship, either, $F(1, 19) = .013$, $p = .911$, which indicated no significant difference in learners' scores for paradigmatic (M = 25.833, SE = .782) and syntagmatic (M = 25.690, SE = .673) relationships. No significant interaction effect between word class and types of association was found, $F(1, 19) = 1.184$, $p = .290$. These findings suggested that word class and type of association essentially had no significant influence on learners' WAT-C scores.

Relationship of Vocabulary Depth with Reading Comprehension

Table 3 shows the bivariate correlations between all the measured competencies. Vocabulary size and depth had a moderate correlation ($r = .620$, $p < .01$), suggesting that they had some overlap but also enough substantial difference that they could be considered to be distinct aspects of vocabulary knowledge. Both vocabulary measures also correlated significantly with both tasks of reading comprehension. Vocabulary size ($r = .748$, $p < .001$) and depth ($r = .720$, $p < .001$) seemed to have a comparable level of correlations with long passage comprehension, whereas for short passage comprehension, the correlation with vocabulary depth ($r = .726$, $p < .001$) appeared to be stronger than that with vocabulary size ($r = .450$, $p < .05$).

Insert Table 3 about here

Two separate sets of hierarchical regression analyses with the long and short passage comprehension tasks as the criterion variables were used to determine the unique contribution of

vocabulary depth to reading comprehension on the one hand and the relative contributions of vocabulary size and depth on the other. As shown in Table 4, in both sets of analyses, orthographic processing and grammatical knowledge were entered into the regression equation first as covariates. This was followed by the two vocabulary knowledge measures. To examine the relative contributions of vocabulary size and depth, the order of their entry into the regression equations was switched. As Table 4 shows, the two covariates together explained about 29.6% and 17.2% of the variance in the long and the short passage comprehension tasks, respectively. Over and above the two covariates, vocabulary size and depth collectively explained about 37.7% and 48.1% of the additional variance in long and short passage comprehension, respectively.

Insert Table 4 about here

While the above result clearly pointed to the importance of vocabulary knowledge in reading comprehension, there were variations in the individual contributions of the two dimensions for the two comprehension tasks. Specifically, vocabulary size explained about 28.1% of the additional variance in long passage comprehension over and above the covariates (p = .004). After size was already entered into the regression model, vocabulary depth still explained about 9.6% of the additional variance (p = .045). Conversely, when vocabulary depth was entered before size, it explained about 24.5% of the additional variance in long passage comprehension (p = .008). After accounting for the effect of depth, vocabulary size still uniquely explained about 13.2% of the additional variance in long passage comprehension (p = .022).

Different patterns, however, were observed when short passage comprehension was the criterion variable. Specifically, when vocabulary size was entered right after the two covariates,

it explained about 10.4% of the additional variance in short passage comprehension, but this effect was not significant ($p = .136$). After controlling for vocabulary size, vocabulary depth explained about 37.7% of the unique variance in short passage comprehension ($p < .001$). When the order of entry was switched, that is, when vocabulary depth was entered before size, it explained about 47.8% of the additional variance in short passage comprehension ($p < .001$). Over and above depth, vocabulary size, however, did not have a significant effect ($p = .725$), and the unique proportion of variance explained in short passage comprehension was negligible (about 0.3%).

Discussion

Effects of Word Frequency on WAT-C

In sum, there were no significant correlations between the word frequencies of the WAT-C items and learners' WAT-C scores, indicating that word frequency did not have a significant impact on depth of vocabulary knowledge. Initially, this result seemed surprising as it would be expected that the more frequent words were, the stronger network learners were likely to build for those words in their mental lexicon because of their cumulative experiences with those words in the process of L2 learning and use. This result also seemed to contradict the finding of Greidahus and Nienhuis (2001), where the frequency of target words significantly discriminated among the performance of Dutch learners of French.

It is conjectured that the divergence of findings might be related to the different frequency profiles of words used in the studies. In Greidahus and Nienhuis's (2001) study, the authors intentionally manipulated the frequency levels of their target words with a clear hierarchy of frequency bands (i.e., from the most frequent 1000 to 5000 words). The words that were chosen for this study, however, all came from the most frequent 2000 words for entry-level

Chinese learners (i.e., Level 1 wordlist in the GCSCW-TCSOL; State Language Commission, 2011). In other words, compared to those in Greidahus and Nienhuis (2001), the words that were targeted in this study might have been so much more frequently encountered by learners that they could not generate any significant frequency effect, particularly for these advanced learners who had had a significant number of years of studying Chinese, in particular at least a year of studying abroad in China. Note that in Greidahus and Nienhuis's (2001) study, despite the many significant effects of word frequency revealed, no significant difference was actually observed between the two lowest frequency levels (i.e., 1000 and 2000) for Level 4 learners (i.e., those who had studied French for 3-4 years at their respective university). In this respect, the findings of that study and this one seemed to converge for very frequent words. The findings also suggest that advanced L2 learners' network knowledge for very frequent words might be so strong that it is not sensitive to the actual level of word frequency (i.e., a "threshold" of word frequency for a frequency effect to occur or not occur).

Effects of Word Class and Type of Association on WAT-C

It was also found that neither word class nor type of association had a significant effect on the WAT-C. Specifically, the learners' performance showed no significant difference between adjectives and verbs on the one hand and paradigmatic and syntagmatic association on the other. Previous studies on English L2 learners using free association tasks generally documented a predominance of paradigmatic over syntagmatic associates in adult L2 learners' mental lexicon (e.g., Nissen & Henriksen, 2006; Wolter, 2001). Jiang (2002) reported a similar finding for Chinese L2 learners. Nissen and Henriksen (2006) further noted that the relative presence of different types of associates in learners' responses may be moderated by the word class of stimulus words. These findings based on free association tasks appeared to suggest that the

learners in this study should have scored higher for paradigmatic association than syntagmatic association.

An explanation for the discrepant findings might be the different nature of free association tasks and the WAT-C, the latter of which only required learners to respond by discerning given associates and distractors. Based on native speakers' canonical responses, Henriksen (2008) calculated Danish-speaking English learners' free association scores and found that those scores were not significantly correlated with the learners' performance on a word connection task (a variant of the word associates test; Read, 1993, 1998). This finding tended to provide support for the author's hypothesis that the two tasks might tap different types of relational knowledge with reference to a hierarchical conceptualization for the mental representation of different types of lexical knowledge in semantic memory. If distinct levels of relational knowledge are drawn upon by learners, it would then make sense that productive preference over paradigmatic associates might not necessarily be translated into a greater score for this type of association in the context of a controlled, recognition task. Such an explanation might also be applicable to the finding reported here – that there was no significant difference between adjectives and verbs, which seems counterintuitive to what would be predicted based on Nissen and Henriksen's (2006) finding that adjectives triggered more paradigmatic as well as syntagmatic associates than verbs. On the other hand, if this explanation holds, why did Greidahus and Nienhuis (2001), which also used a controlled recognition task, find significantly better scores for paradigmatic association than syntagmatic association? The specific reason was not clear, although it might be related to Greidahus and Nienhuis's (2011) lack of attention to word frequency as another independent variable when the scores for paradigmatic and syntagmatic associates were compared.

Vocabulary Depth and Size and Reading Comprehension

Vocabulary depth as measured by the WAT-C contributed uniquely to reading comprehension as a distinct aspect of vocabulary knowledge, and the relative contributions of vocabulary size and depth varied across the two reading comprehension tasks. Specifically, for short passage comprehension, vocabulary depth was a significant and unique predictor of reading comprehension, and it was a much stronger predictor than size; on the other hand, for the long passage comprehension task, both depth and size were significant and unique predictors, but size was a stronger predictor.

The unique effect of vocabulary depth on reading comprehension for both tasks was not a surprise, as the comprehension process involves not only knowledge of basic meanings of those words that appear in a text, but also how those words are related. Such a finding also agrees with previous findings on L2 learners of English (e.g., Qian, 1999, 2002). The findings that vocabulary depth, as opposed to vocabulary size, was a unique predictor of short passage comprehension and that vocabulary size was a more important predictor of long passage comprehension might be related to the variations in the textual properties and the comprehension questions between the two tasks.

Specifically, the short passages were overall lexically very simple and the words there might have been largely known to all the participants. As a result, the variance in short passage comprehension performance might primarily reflect the learners' knowledge of meaning relationships (i.e., depth of vocabulary knowledge as measured by the WAT-C) rather than the actual meanings of the words in the texts where minimal individual difference might have existed. In addition, the questions for the short passage comprehension task focused exclusively on learners' inferencing skills, which required deeper cognitive processing of words and their

meaning relationships. More specifically, a strong knowledge of how words in the passages were related to others words that were not present would be essential for successful textual inferencing and construction of a situation model (Kintsch, 1998). In this respect, the assessment focus of the short passage comprehension task appeared to show a similarity in cognitive demands as the summary writing task in Li and Kirby (2015) where English learners' vocabulary depth was found to play a more important role than size.

In contrast, the long passages were much more complex lexically, that is, they had a much greater presence of low frequency words, and the comprehension questions were primarily literal that asked learners to attend to explicit information. It thus seemed reasonable to expect that the task would require greater involvement of learners' knowledge of meanings of words in the stories, and consequently, vocabulary size emerged as a more important predictor than depth. The finding that vocabulary size was a stronger predictor of long passage comprehension should, of course, not be wrongly interpreted to mean that vocabulary depth is unimportant to comprehending complex texts and answering questions that test literal comprehension. After all, literal comprehension also requires learners to be able to integrate word meanings for the construction of a text model, which is obviously related to learners' knowledge about the meaning relations of words presented in a text.

Pedagogically, the findings of this study suggest that while it is important for learners to develop an understanding of basic meanings for a large number of words, it is also important for them to build and consolidate a network of how the meanings of those words are related and organized in the mental lexicon. Some analyses of L2 textbooks, albeit not Chinese ones, however, revealed that the two aspects of vocabulary knowledge that received the most attention were often word forms and meanings (for establishing form-meaning connections) with little

coverage of critical aspects of vocabulary depth like associations (Brown, 2011; Neary-Sundquist, 2015). While the vocabulary focus of Chinese textbooks is unknown, there have been constant reports that Chinese learners, even advanced ones, tended to show a lack of fine-grained understandings of word meanings, such as nuanced meanings of synonyms and their distinct collocation patterns (e.g., 增进[理解] / *deepen* or *promote* [*understanding*] versus 增加[知识] / *improve* or *broaden* [knowledge]) (e.g., Li, 2016; Xing, 2013). Yet, instructionally, because of the character-based writing system, instructional attention has been primarily given to learners' ability to recognize printed words, which are typically composed of multiple characters (Everson, 2011; Shen, 2013). Thus, while initial form-meaning connection is an important first step (e.g., character recognition in Chinese), the goal of vocabulary instruction should go far beyond it (Schmitt, 2008). As Schmitt (2008) argued, vocabulary instruction needs to create opportunities for learners to have "engagement" with words (i.e., "more exposure, attention, manipulation, or time spent on lexical items") (p. 339). For example, classroom activities can be organized to engage learners to analyze semantic features of words over and beyond recognition of synonyms and antonyms presented in graphic organizers, and to promote their use of lexical collocations (i.e., chunking-based lexical learning) based on those analyses beyond the memorization of glossary lists in the textbook. In addition to activities to promote the building of a stronger network as well as other aspects of word knowledge (Haastrup & Henriksen, 2000), it should also be important for teachers to engage learners to attend to their own learning process and use appropriate strategies for self-regulated learning of words (Schmitt, 2008; Shen & Xu, 2015; Tseng & Schmitt, 2008). In this regard, many of the strategies, methods, and activities (SMAs) that Shen and Xu (2015) highlighted in their study on an Active Learning approach to Chinese

vocabulary instruction are useful, such as presenting learners with clear learning goals and encouraging self-discoveries of lexical patterns.

The unique contribution of vocabulary depth to reading comprehension over and above vocabulary size reinforces the aforementioned emphasis on the importance of building strong network knowledge. In other words, knowing the basic meanings of the words that appear in a text, an issue that concerns lexical coverage, would not result in an enhanced level of comprehension if word relationships are unclear to learners. With particular respect to the importance of vocabulary depth for deep comprehension like textual inferencing, it would be desirable for instruction to go beyond the basic meanings of target vocabulary words to include their culturally-loaded (e.g., connotation) and pragmatic meanings as well. In addition, it would also be a useful strategy to group words thematically to help learners establish schemata and activate them for the comprehension of texts where those words are used.

Limitations

When seeking to generalize the findings of this study, several limitations should be considered. Firstly, the frequency bands of the words in the WAT-C were not manipulated, primarily because we were concerned that low frequency words or words that were unfamiliar to learners might lead to guessing and consequently would threaten the validity of the test (Read, 2004; Schmitt et al., 2011). Thus, all the target and choice words were highly frequent and defined as essential for entry level Chinese learners (State Language Commission, 2011). Nevertheless, in the future, it would be interesting to include words from distinct frequency bands to explore any possible word frequency effects on Chinese learners' association knowledge. Secondly, when the interaction between reader (i.e., variabilities in different types of vocabulary knowledge) and non-reader variables (RAND Reading Group, 2002) in reading comprehension was addressed,

complex manipulations of the type of text (e.g., genre) and the type of questions (e.g., question format in addition to measuring different types of comprehension) were not considered: The short and long passages were almost all narrative texts and the questions for the short passages focused only on inferencing, whereas those for the long passages focused primarily on literal comprehension. Thus, it was not possible to obtain a more nuanced understanding of the reader-text-question/task interaction that has been revealed in previous studies (e.g., Eason et al., 2012; Li & Kirby, 2014). This issue certainly deserves more attention in future research on Chinese reading comprehension and L2 reading comprehension in general. Lastly, because the sample size was small and all participants had a relatively high level of proficiency in Chinese as a second language, it was impossible for us to identify possible differences in performance and predictive relationships between higher and lower proficiency learners.

Conclusion

Using a Chinese WAT, this study examined Chinese L2 learners' depth of vocabulary knowledge and the contribution of this type of knowledge to reading comprehension. No significant effect of word frequency on learners' WAT-C performance was found. In addition, learners' performance for adjectives and verbs and for paradigmatic and syntagmatic associations were similar. More importantly, vocabulary depth was found to be a significant and unique predictor of reading comprehension over and above vocabulary size, disregarding the types of texts that were read and the comprehension skills that were assessed. However, the relative contributions of vocabulary depth and size varied between the short and long passage comprehension tasks. While there are many more issues of L2 Chinese learners' vocabulary depth that may be explored, the findings of the present study clearly highlight its impact on L2 Chinese learners' reading comprehension.

Acknowledgement

The authors would like to thank Dr. Anne Nerenz and the two anonymous reviewers for their insightful and constructive comments on an earlier version of this paper. Any error is ours.

References

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77-117). Newark, DE: International Reading Association.

Beijing Language Institute. (1986). *Modern Chinese frequency dictionary*. Beijing: Beijing Language Institute Press.

Brown, D. (2011). What aspects of vocabulary knowledge do textbooks give attention to? *Language Teaching Research, 15*, 93-97.

Eason, S. H., Goldberg, L. F., Young, K. M., Geist, M. C., & Cutting, L. E. (2012). Reader-text interactions: How differential text and question types influence cognitive skills needed for reading comprehension. *Journal of Educational Psychology, 104*, 515-528.

Everson, M. (2011). Best practices in teaching logographic and non-roman writing systems to L2 learners. *Annual Review of Applied Linguistics, 31*, 249-274.

Fitzpatrick, T. (2013). Word associations. In C. Chapelle (Ed.), *Encyclopedia of applied linguistics* (pp. 6193-6199*)*. Oxford: Wiley-Blackwell.

Fitzpatrick, T., Playfoot, D., Wray, A., & Wright, M. (2015). Establishing the reliability of word association data for investigating individual and group differences. *Applied linguistics, 36*, 23-50.

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. NY: Cambridge University Press.

Greidanus, T., Bogaards, P., van der Linden, E., Nienhuis, L., & de Wolf, T. (2004). In P. Bogaards, & B. Laufer-Dvorkin (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 191-208). Amsterdam: John Benjamins.

Greidanus, T., & Nienhus, L. (2001). Testing the quality of word knowledge in a second language by means of word associations: Types of distractors and types of associations. *The Modern Language Journal, 84*, 567-577.

Haastrup, K. & Henriksen, B. (2000). Vocabulary acquisition: Acquiring depth of knowledge through network building. *International Journal of Applied Linguistics, 10*, 221-240.

Hamilton, S. T., Freed, E. M., & Long, D. L. (2013). Modeling reader and text interactions during narrative comprehension: A test of the Lexical Quality Hypothesis. *Discourse Processes, 50*, 139-163.

Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition, 21*, 303-317.

Henriksen, B. (2008). Declarative lexical knowledge. In D. Albrechtsen, K. Haastrup, & B. Henriksen (Eds.), *Vocabulary and writing in a first and second language: Processes and development* (pp. 22-66). New York: Palgrave Macmillan.

Horiba, Y. (2012). Word knowledge and its relationship to text comprehension: A comparative study of Chinese- and Korean-speaking L2 learners and L1 speakers of Japanese. *The Modern Languages Journal, 96*, 108-121.

Hu, M. H.-C., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*, 403-430.

Jeon, E. H. & Yamashita, J. (2014). L2 reading comprehension and its correlated: A meta-analysis. *Language Learning, 64*, 160-212.

Jiang, S. (2002). Chinese word associations for English speaking learners of Chinese as a Second Language. *Journal of the Chinese Language Teachers Association, 37*(3), 55-70.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge: Cambridge University Press.

Koda, K. (2005). *Insights into second language reading*. Cambridge: Cambridge University Press.

Li, M., & Kirby, J. R. (2015). The effects of vocabulary breadth and depth on English reading. *Applied Linguistics*, 36, 611-634.

Li, S. (2016). A corpus-based analysis of collocational errors by American learners of Chinese and its implications for the teaching of vocabulary. *Chinese as a Second Language, 51*, 62-78.

Meara, P. (1980). Vocabulary acquisition: A neglected aspect of language learning. *Language Teaching, 13*, 221-246.

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.

Milton, J. (2013). Measuring the contribution of vocabulary knowledge to proficiency in the four skills. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis* (pp. 57-78). Euro SLA.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, P., 2006. How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review, 63*, pp.59–82.

Neary-Sundquist, C. A. (2015). Aspects of vocabulary knowledge in German textbooks. *Foreign Language Annals, 48*, 68-81.

Nelson, K. (1977). The syntagmatic-paradigmatic shift revisited: A review of research and theory. *Psychological Bulletin, 84*, 93–116.

Nissen, H., & Henriksen, B. (2006). Word class influence on word association test results. *International Journal of Applied Linguistics, 16*, 389-408.

RAND Reading Study Group. (2002). *Reading for understanding: Toward a research and development program in reading comprehension.* Santa Monica, CA: Office of Education Research and Improvement.

Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review, 56*, 282-307.

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning, 52*, 513-536.

Qian, D. D., & Schedl, M. (2004). Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing, 21*, 28-52.

Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly, 10*, 77-89.

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing, 10*, 355-371.

Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A. J. Kunnan (Ed.), *Validation in language assessment* (pp. 41-60). Mahwah, NJ: Lawrence Erlbaum Associates.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined. In B. Laufer & P. Bogaards (Eds.), *Vocabulary in a second language: Selection, acquisition and testing* (pp. 209-227). Amsterdam: John Benjamins.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching, 12*, 329-363.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. New York: Palgrave Macmillan.

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning, 64*, 913-951.

Schmitt, N., Ng, J. W. C., & Garras, J. (2011). The Word Associates Formats: Validation evidence. *Language Testing, 28*, 105-126.

Schoonen, R., & Verhallen, M. (2008). The assessment of deep word knowledge in young first and second language learners. *Language Testing, 25*, 211-236.

Shen, H. (2013). Chinese L2 literacy development: Cognitive characteristics, learning strategies, and pedagogical interventions. *Language and Linguistic Compass, 7*, 371-387.

Shen, H., &Xu, W. (2015). Active learning: Qualitative inquiries into vocabulary instruction in Chinese L2 classrooms. *Foreign Language Annals, 48*, 82-99.

State Language Commission. (2011). *Graded Chinese syllables, characters and words for the application of Teaching Chinese to Speakers of Other Languages.* Beijing: State Language Commission, Chinese Ministry of Education.

Tseng, W. & Schmitt, N., 2008. Toward a Model of Motivated Vocabulary Learning: A Structural Equation Modeling Approach. *Language Learning*, 58, 357-400.

Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics, 22*, 217-234.

Urquhart, A. H., & Weir, C. J. (1998). *Reading in a second language: Process, product and practice*. London: Longman.

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth vs. breadth. *Canadian Modern Language Review, 53*, 13-39.

Wolter, B. (2001). *Comparing the L1 and L2 mental lexicon. Studies in Second Language Acquisition, 23*, 41–69.

Xing, H. (2013). Collocation knowledge and second language lexical acquisition (in Chinese). *Yuyan Yingyong Yanjiu (Applied Linguistics), 21*(4), 117-126.

Zhang, D. (2012). Vocabulary and grammar knowledge in L2 reading comprehension: A structural equation modeling study. *Modern Language Journal, 96*, 554-571.

Table 1. *Raw and Log10 Frequency of WAT-C Words*

		Adjective: M(SD)	Verbs: M(SD)	Total: M(SD)
Target Words	Number of Items	10	10	20
	Raw	693.20(535.84)	1780.10(1474.12)	1236.65(1215.00)
	Log	2.69(0.42)	3.13(0.33)	2.91(0.43)
Choice Words	Number of Items	30	30	60
	Raw Paradigmatic	1293.10(1287.86)	1514.40(1081.45)	1403.75(1162.98)
	Log Paradigmatic	2.69(0.29)	2.92(0.32)	2.81(0.32)
	Number of Items	30	30	60
	Raw Syntagmatic	1145.50(761.62)	2432.23(1935.72)	1788.87(1576.51)
	Log Syntagmatic	2.70(0.32)	3.12(0.38)	2.91(0.40)
	Number of Items	60	60	120
	Raw Total	1219.30(934.41)	1973.32(1052.07)	1596.31(1042.83)
	Log Total	2.70(0.26)	3.02(0.25)	286(0.30)
Total	Number of Items	70	70	140
	Raw	1144.14(839.13)	1907.51(940.23)	1525.83(951.66)
	Log	2.70(0.25)	3.02(0.23)	2.86(0.29)

Table 2. *Means, Standard Deviations, and Reliability of Tasks*

	n	*Mean*	*SD*	Reliability (α)
Orthographic Processing	20	18.52	0.93	.607
Grammaticality Judgement	20	16.86	2.13	.655
Vocabulary Size	30	20.71	6.03	.942
Vocabulary Depth	20	103.05	13.25	.879
Adjectives	10	51.29	5.95	.762
- Paradigmatic Relations	-	26.10	2.96	-
- Syntagmatic Relations	-	25.19	3.46	-
Verbs	10	51.76	7.78	.844
- Paradigmatic Relations	-	25.57	5.09	-
- Syntagmatic Relations	-	26.19	3.53	-
Short Passage Comprehension	15	12.62	1.43	.861
Long Passage Comprehension	20	8.71	4.31	.825

Table 3. *Bivariate Correlations Between All Measure Variables*

		1	2	3	4	5	6
1	Orthographic Processing	–					
2	Grammaticality Judgement	0.343	–				
3	Vocabulary Size	0.385	.557***	–			
4	Vocabulary Depth	.441*	.570**	.620**	–		
5	Short Passage Comprehension	0.007	0.392	.450*	.726***	–	
6	Long Passage Comprehension	0.301	.529*	.748***	.720***	.540*	–

* $p < .05$ ** $p < .01$ *** $p < .001$

Table 4. *Hierarchical Regression Analysis Testing the Relative Contributions of Vocabulary Size and Depth to Reading Comprehension*

Step	Predictor	Long Passage Comprehension				Short Passage Comprehension			
		R^2	ΔR^2	ΔF	p	R^2	ΔR^2	ΔF	p
1	Orthographic Processing	0.091	0.091	1.898	0.184	0.000	0.000	0.001	0.975
2	Grammaticality Judgement	0.296	0.205	5.249	0.034	0.172	0.172	3.734	0.069
3	Vocabulary Size	0.577	0.281	11.310	0.004	0.276	0.104	2.447	0.136
4	Vocabulary Depth	0.673	0.096	4.705	0.045	0.653	0.377	17.344	0.001
3	Vocabulary Depth	0.541	0.245	9.073	0.008	0.65	0.478	23.205	0.000
4	Vocabulary Size	0.673	0.132	6.481	0.022	0.653	0.003	0.128	0.725